Spotify:
Stream and Sound Analysis Over Time

Alex Arnold
Scott Frazier
Samantha Lane
Jenette Smith

Introduction:

Modest Mouse said it best, "Music is to the soul what words are to the mind."[1] Music helps you think, feel, and express. We were inspired to investigate Spotify and music streaming after finding really interesting visualizations from Tableau's public gallery investigating top music tracks for 2020.

So what music do we listen to? Has the music we listen to changed over time? Who listens the most? This report will explore which countries stream the most music and whether or not sounds by audio feature measurement have changed over time through analysis using interactive visualizations.

Research Questions:

- What countries listened to/streamed the most music?
- What are the top countries that listened to?
- How does music sound change over time?
- How is sound broken down and is it correlated?

Hypotheses:

We believe that the United States streams the most music and that audio features, such as speechiness and danceability, have increased over time.

Data Sources:

We collected data from a variety of sources. The bulk of our project used a kaggle data set that consisted of over 160,000 songs from the years 1921 to now. The data set itself consisted of five CSVs that contained information regarding each track.

We primarily used the CSV that had split the tracks up by the year they were recorded. We also created our own CSV by pulling the weekly top 100 for each week of 2020 from the Spotify website. Lastly, we used a CSV from a Tableau graph that included information regarding streams per country around the world in combination with a census CSV to build a choropleth map of the world.

Analysis:

We created a website to present two visual dashboards to further investigate our hypotheses and divided these dashboards by streams and sound.

[1] https://graciousquotes.com/music-quotes/

Our group used bootstrap on this entire website project. By utilizing bootstrap, we are able to modify the CSS and JavaScript code to create a fully responsive and interactive website. Bootstrap also helps with how a website is displayed from a mobile platform or resizing to a computer monitor. We also used Bootswatch CSS to further enhance the styling, page layouts, etc. in order to adapt our code enhancing our project.

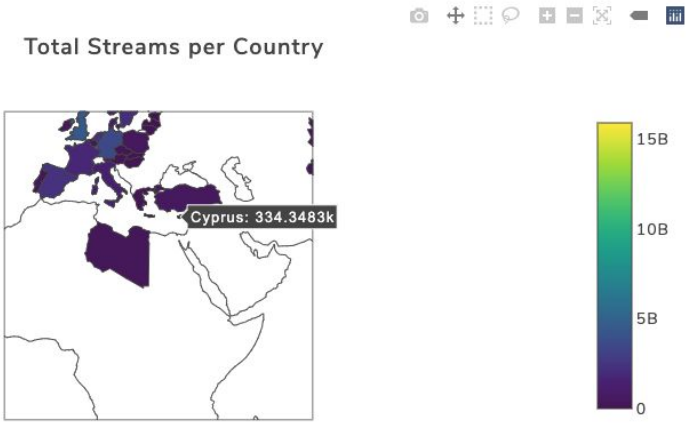
Streams Dashboard

Our streams dashboard highlights the number of streams by country as well as the top streamed songs in 2020.

World Map of Countries by Number of Streams:

Using plotly in combination with the Tableau dataset and Census CSV, we were able to create a choropleth globe that is colored by the number of streams for each country.

You are able to drag the globe to rotate towards different countries. As you hover over the map, the name of the country and the number of streams for that country are displayed. The color scale is represented on the right hand side of the graphic and shows that the countries with higher stream counts are a yellow hue and countries with smaller stream countries are a dark purple hue.

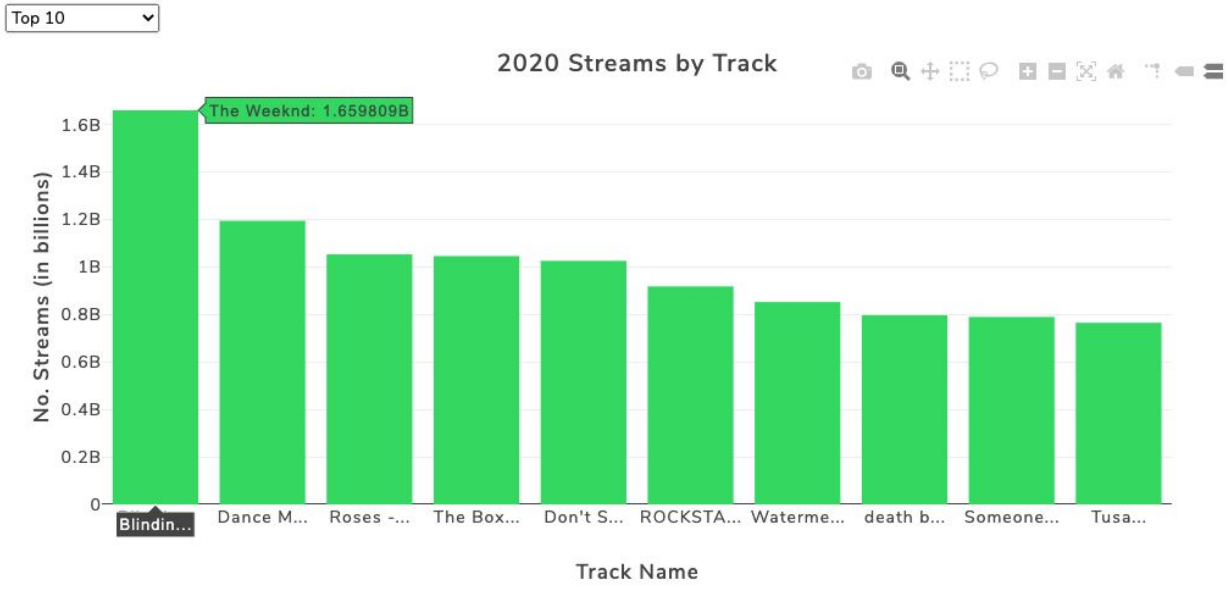


The image above shows Cyprus, the country with the lowest number of streams. By hovering over the country you can see the value of streams which is about 300,000.

Next you can see the image of the country with the highest number of streams, The United States with about 16 billion streams.

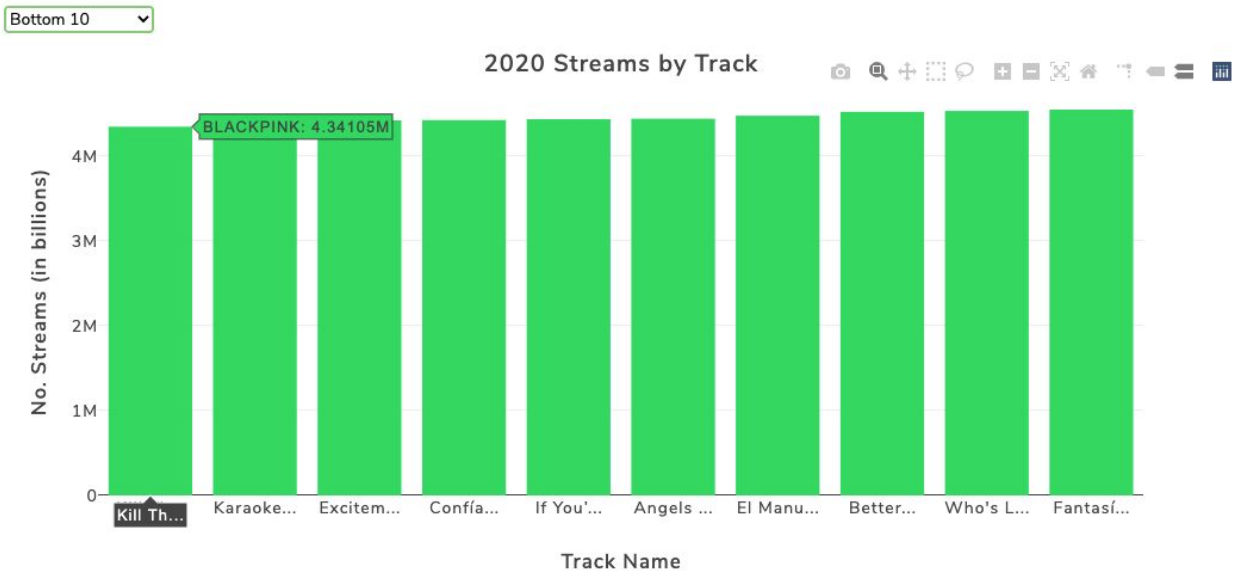
Streams of 2020:

Using plotly we created a bar chart of the Top 100 Tracks week over week in 2020 that can be filtered based off of the Top 10, Top 20, Bottom 10, and Bottom 20 to show which songs were streamed the most and which songs were streamed the least in the Top 100.



The bar chart above shows the Top 10 songs in descending order by number of streams. You can see the name of the artist, number of streams and insight into the song title by hovering

over each bar. “Blinding Lights” by The Weeknd was the most streamed song of 2020 with a total of about 1.7 billion streams.



This bar chart shows the Bottom 10 songs in ascending order by number of streams. In conclusion, “Kill This Love” by BLACKPINK was the least streamed song out of the Top 100 week over week in 2020.

Sound Dashboard

Audio Features:

Our sound dashboard highlights the audio features of the tracks we investigated from our Kaggle data set. Audio features are the characteristics of a track determined by Spotify and listed for reference in the Spotify API. These features provide additional insight into each track.

- **Acousticness-** confidence measure of whether the track is acoustic
- **Danceability-** how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity
- **Energy-** represents a perceptual measure of intensity and activity
- **Instrumentalness-** predicts whether a track contains vocals
- **Liveness-** detects the presence of an audience in the recording
- **Speechiness**- detects presence of spoken words in a track
- **Valence-** describes musical positiveness conveyed by a track

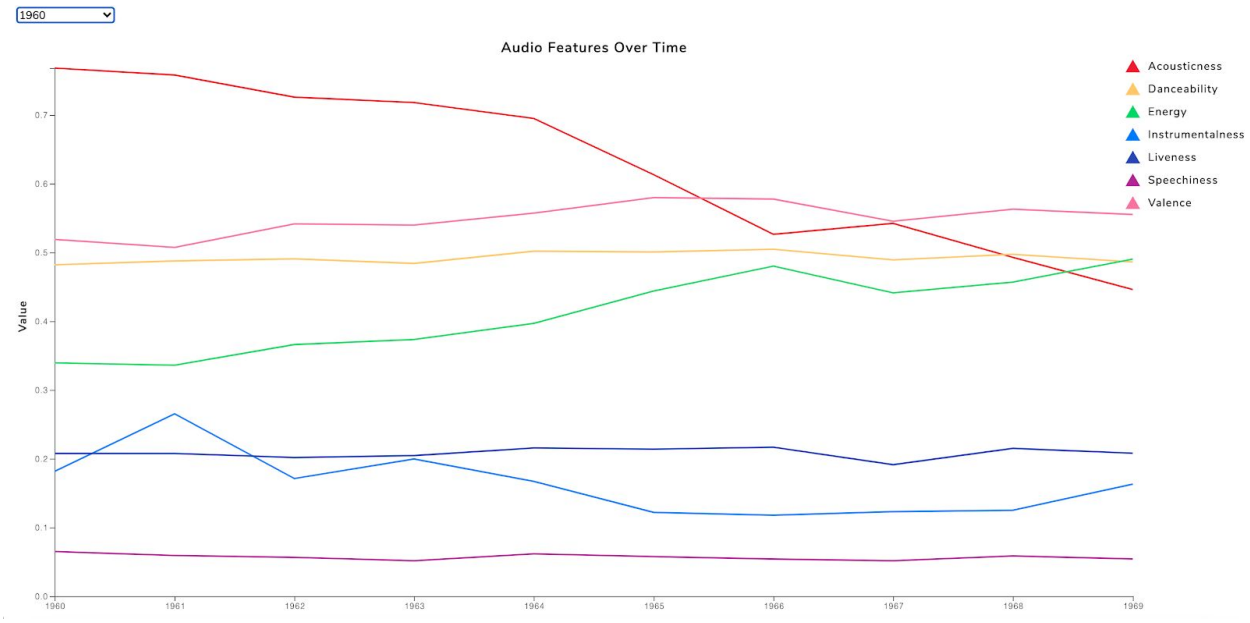
The above audio features are valued on a scale between 0 and 1, in which 1 represents a high confidence for that feature. In addition, the below features are measured on a scale dependent on that feature.

- **Key-** Primary key of track
- **Mode-** indicates the modality (major or minor) of a track
- **Duration_ms-** length of track in milliseconds
- **Loudness-** overall loudness of a track in decibels (dB)
- **Popularity**- popularity of an album calculated by the individual tracks (scored 0 to 100-most popular)
- **Tempo-** estimated tempo of track in beats per minutes (BPM)

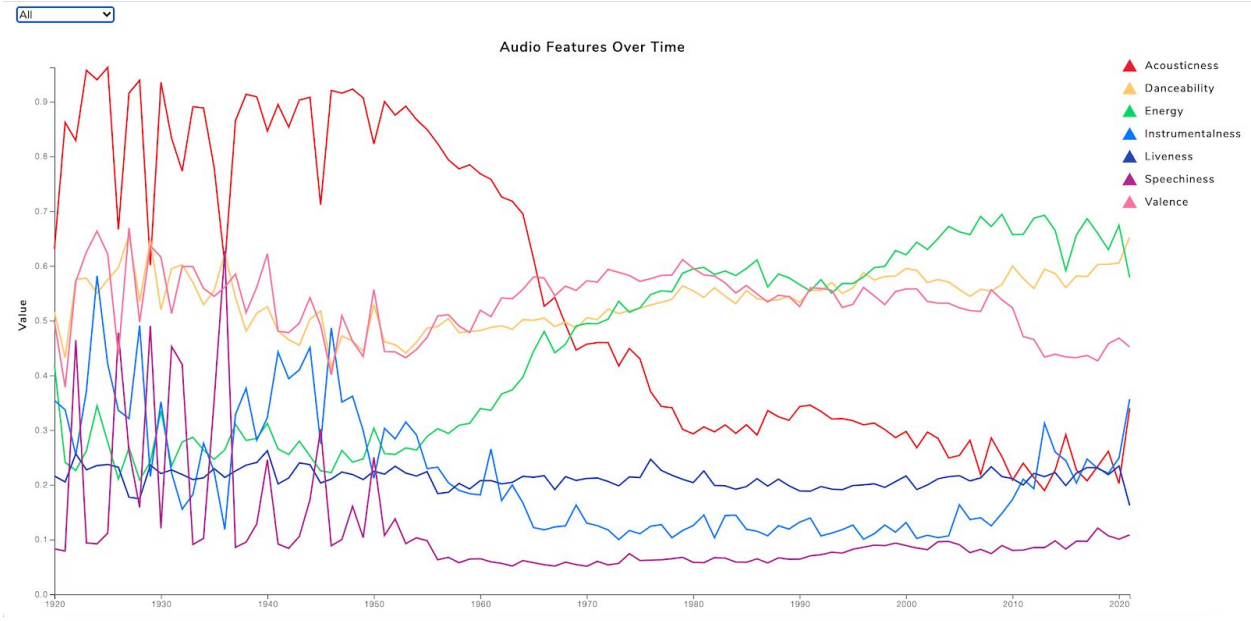
We investigated a variety of these features further in our Sound Dashboard.

Change in Audio Features Over Time:

Using d3, we created a line chart that displays the audio features of Acousticness, Danceability, Energy, Instrumentalness, Liveness, Speechiness, and Valance over time. You are able to filter the data to show the entire time period 1920-2021, or filter to drill down the data by decade.



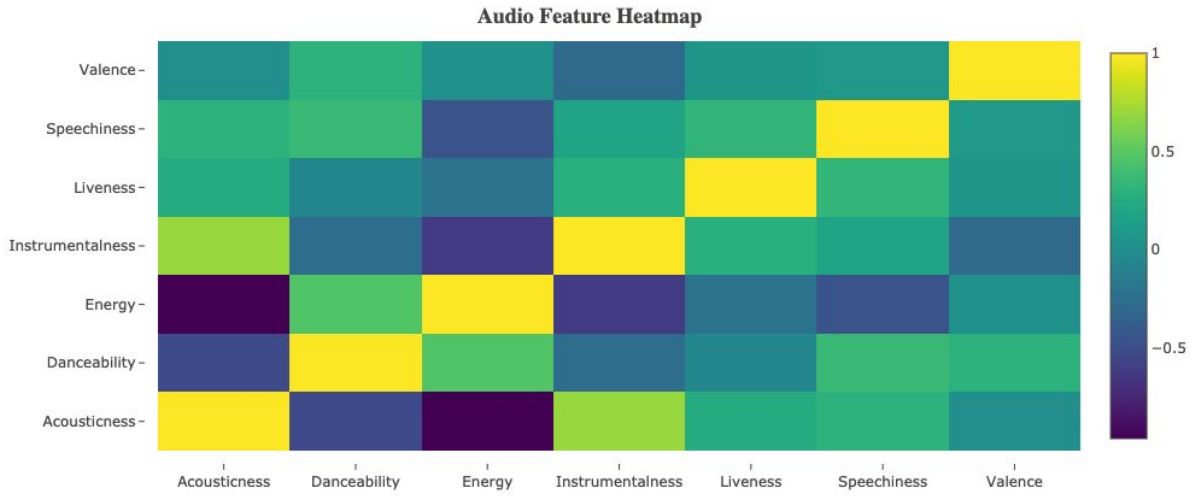
Acousticness and speechiness decrease over time. Danceability, energy increase over time. Valence and Liveness remains relatively the same as does Instrumentalness even though it has significant peaks and valleys within certain decades.



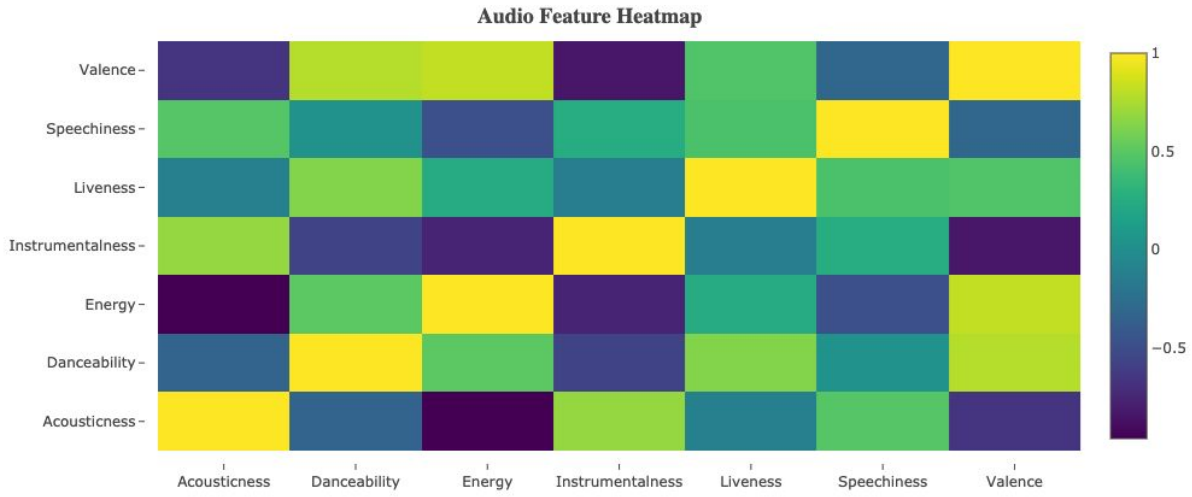
From the line chart, we can determine that sound as it relates to these audio features does change over time. For example, we see a drop in Acousticness between 1960 and 1980 as further shown in the filtered line chart.

Audio Feature Correlations:

We created a correlation heatmap that displays the relationships between the audio features themselves using plotly.



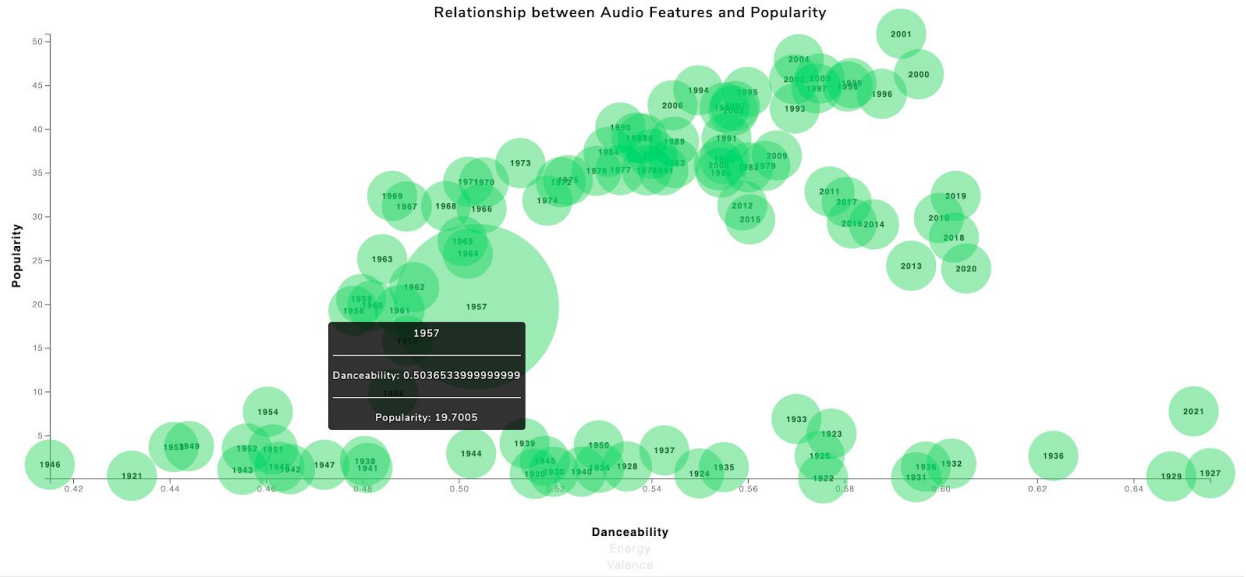
The above heatmap displays the correlations over the entire period in which we can see that Instrumentalness and Acousticness have a strong correlation at 0.69 while Energy and Acousticness have a weak correlation at -0.96. The heatmap can also be filtered by decade using the same filter as the line map.



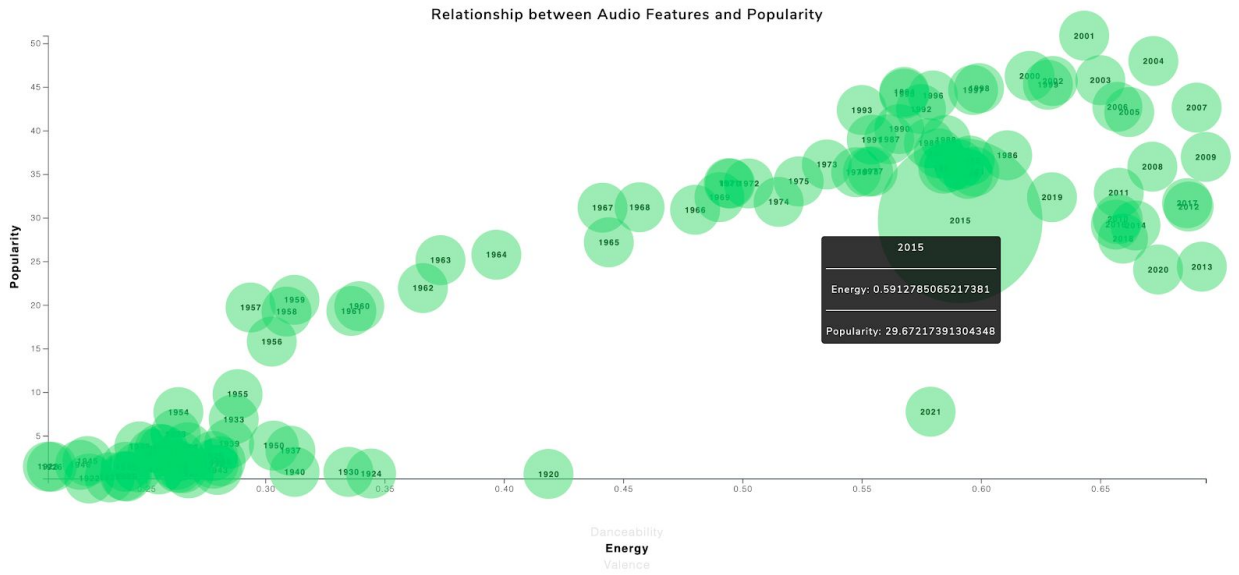
In the above heatmap, we have filtered to the same 1960-1970 time period, further illustrating the changes in relationship of these features at this time. With this time filter, Energy and Valence now have the strongest correlation at 0.82 while Energy and Acousticness still have the weakest correlation at -0.96.

Relationship between Popularity and Audio Features:

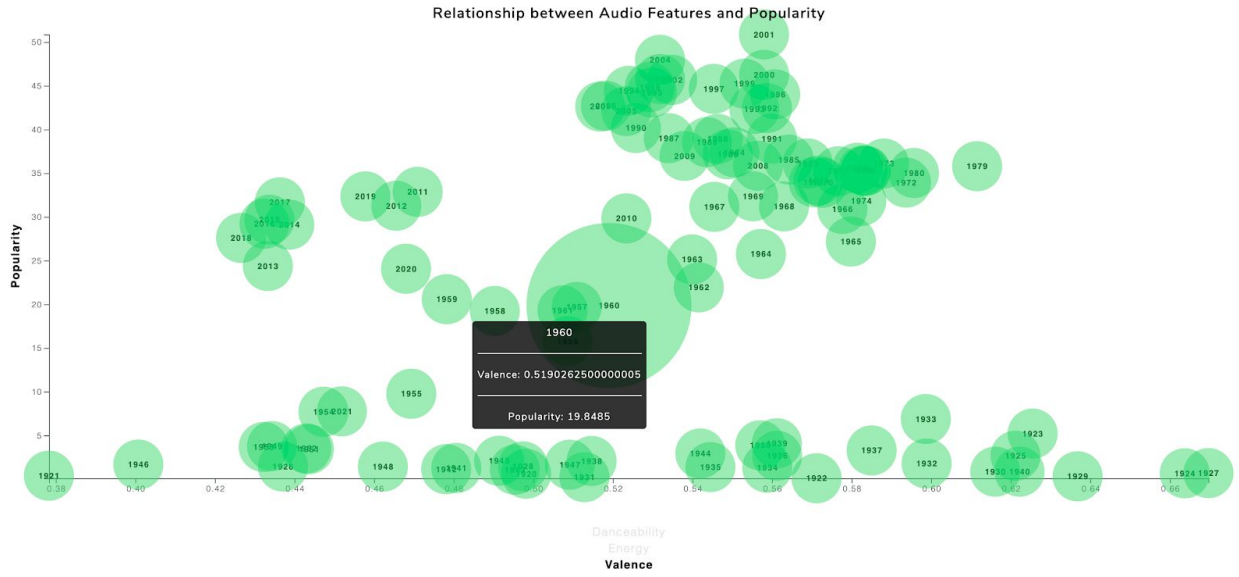
Energy, Danceability, and Valence are the highest valued audio features at the present time. We wanted to investigate those features as they related to the Popularity of a track and created the scatterplot below using d3. The scatterplot is interactive in that the x-axis can be changed to highlight a different audio feature. In addition, each circle represents a year from the dataset and the tooltip highlights that year's popularity score and specific audio feature score.



When analyzing Popularity and Danceability, we can see a slightly positive correlation in the latter part of the 21st century from 1955 to 2020. The year 2021 is currently an outlier as we do not yet have the data over the entire year.



When analyzing Popularity and Energy, there is a strong positive correlation over time. The year 2021 is again an outlier for also not having the data from the entire year.



When analyzing Popularity and Valence, there is no correlation between these audio features over time.

Data Tables:

We also created data tables for our most used data sources to easily filter data for viewing. The first data table was built from one of the Kaggle CSV's that included the year each track was recorded with a filter that allows a year input to display each track for that year. The next data table was built from the Top 100 week over week data set that included a filter for Top 10, Top 20, Bottom 10 and Bottom 20.

BY YEAR

Filter Search

Enter a Year

2010

FILTER TABLE

CLEAR FILTERS

COPY EXCEL CSV PDF

Search:

YEAR	ACOUSTICNESS	DANCEABILITY	DURATION_MS	ENERGY	INSTRUMENTALNESS	LIVENESS	LOUDNESS	SPEECHINESS	TEMPO	VALENCE
1920	0.63	0.52	238093.00	0.42	0.35	0.22	-12.65	0.08	113.23	0.50
1921	0.86	0.43	257891.76	0.24	0.34	0.21	-16.81	0.08	102.43	0.38
1922	0.83	0.58	140135.14	0.23	0.25	0.26	-20.84	0.46	100.03	0.57
1923	0.96	0.58	177942.36	0.26	0.37	0.23	-14.13	0.09	114.01	0.63
1924	0.94	0.55	191046.71	0.34	0.58	0.24	-14.23	0.09	120.69	0.66
1925	0.96	0.57	184777.62	0.28	0.42	0.24	-14.17	0.11	115.48	0.62
1926	0.67	0.60	157836.35	0.21	0.34	0.23	-18.49	0.48	109.57	0.44
1927	0.92	0.66	176542.81	0.27	0.32	0.18	-15.24	0.26	112.66	0.67
1928	0.94	0.53	214396.03	0.21	0.49	0.17	-17.13	0.16	106.75	0.50
1929	0.60	0.65	168795.61	0.24	0.22	0.24	-16.53	0.49	110.93	0.64

BY STREAMS

Filter Search

Select Rankings

All

FILTER TABLE

CLEAR FILTERS

COPY EXCEL CSV PDF

Search:

TRACK NAME	ARTISTS	STREAMS
<3	Bad Bunny	21733626
+Linda	Dalex	19291759
00:00 (Zero O'Clock)	BTS	9372219
10,000 Hours (with Justin Bieber)	Dan + Shay	287267463
1000 Doves	Lady Gaga	6069544
100k Cash	Capital Bra	5192813
11 PM	Maluma	14881517
120	Bad Bunny	10957181
2002	Anne-Marie	9674257
2012	Bausa	4685894

Conclusion:

Our study shows that we accepted the alternative hypothesis in that the United States does stream the most music using Spotify. In terms of sound, we failed to reject the null hypothesis. Danceability has increased over time, but speechiness has not increased over time.

Recommendation:

Based on our analysis, we recommend that artists who want to increase their popularity on Spotify, highlight and increase the audio features of danceability, energy, and valence. These audio feature categories have shown increased growth over time which have led to increases in popularity, assuming there is a relationship between streams and popularity ultimately leading to an increase in the number of streams.

Limitations:

When dealing with music streaming, availability to internet and streaming services, is paramount. Therefore our data is biased towards those who have access to the Internet and ultimately Spotify. When revisiting our choropleth map, not all countries are colored and scaled because of this.

For investigating the streams of countries, our CSV from Tableau was limited to streams through 2017. Therefore, it would have been great to investigate further streaming through 2021. However, given the scope of this project and our time constraints, we focused on this dataset.

The Spotify Kaggle dataset provided genre information but this information was extremely specific and vast, making it difficult to sort in a way to aid in visualizations. Future work investigating genres further while pulling in additional Spotify API data to help would be helpful and extremely interesting.

Our d3 knowledge is limited and provided a challenge when filtering the years with the line charts, yet we persevered. Time was another influencing factor as we can always do more with more time. Yet we are extremely proud of what we accomplished given the extenuating circumstances of the winter storm.

Future Work:

In order to further investigate the vast data that spotify provides it would be vital to use Spotify's API. This would allow us to gather information regarding streams, users, artists, followers, personalization as well as a variety of other different topics. It would also give us the opportunity to get updated and current data.

It would also be beneficial for us to use the updated data from the API to create a MapBox map with the country data provided for each user and GeoJSON data. We would be able to create marker clusters for users around the world and see where the bulk of Spoitfy users are located.

Another idea for the future would be to take into account the population of a country when determining which country had the most streams. In order to see true representation of

streams per person. We could also use user information from the API to determine which country the user represents.

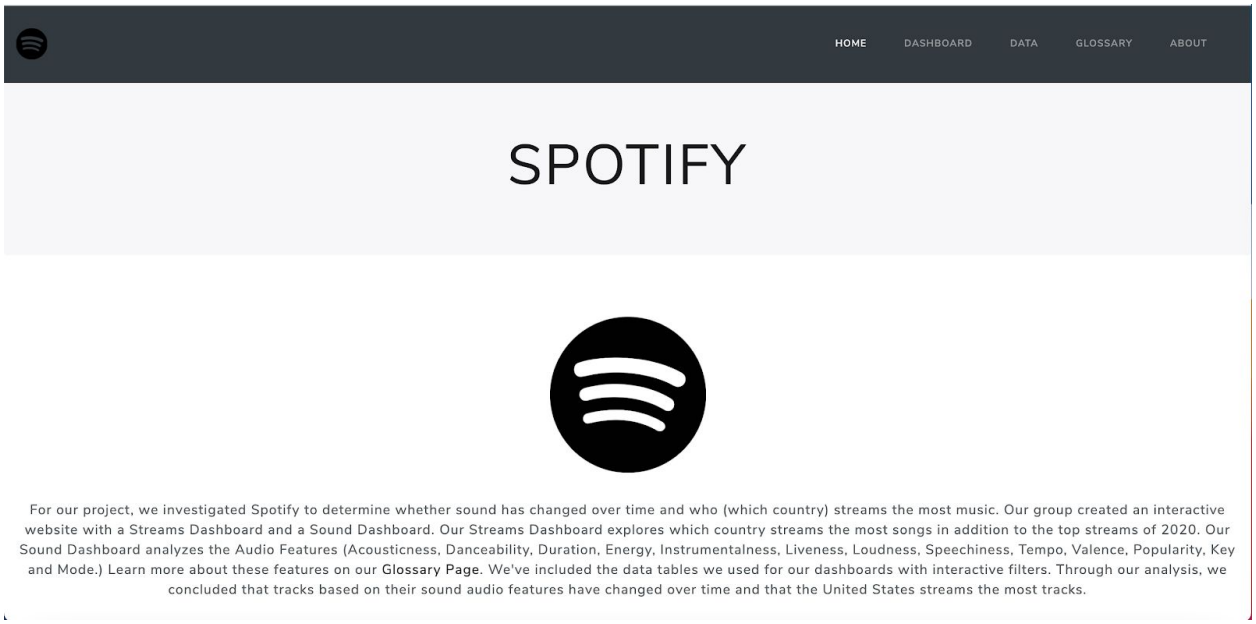
To continue, we would love to investigate genres using the Spotify API. We could compare data between genres and users, artists, tracks, and audio features. It would be interesting to see what genres are most popular, what are the specific audio features of each genre, as well as if artists are consistently in the same genre.

Lastly, it would be very interesting to compare the data provided by Spotify's API to one of their competitors such as Apple Music, who also has an API.

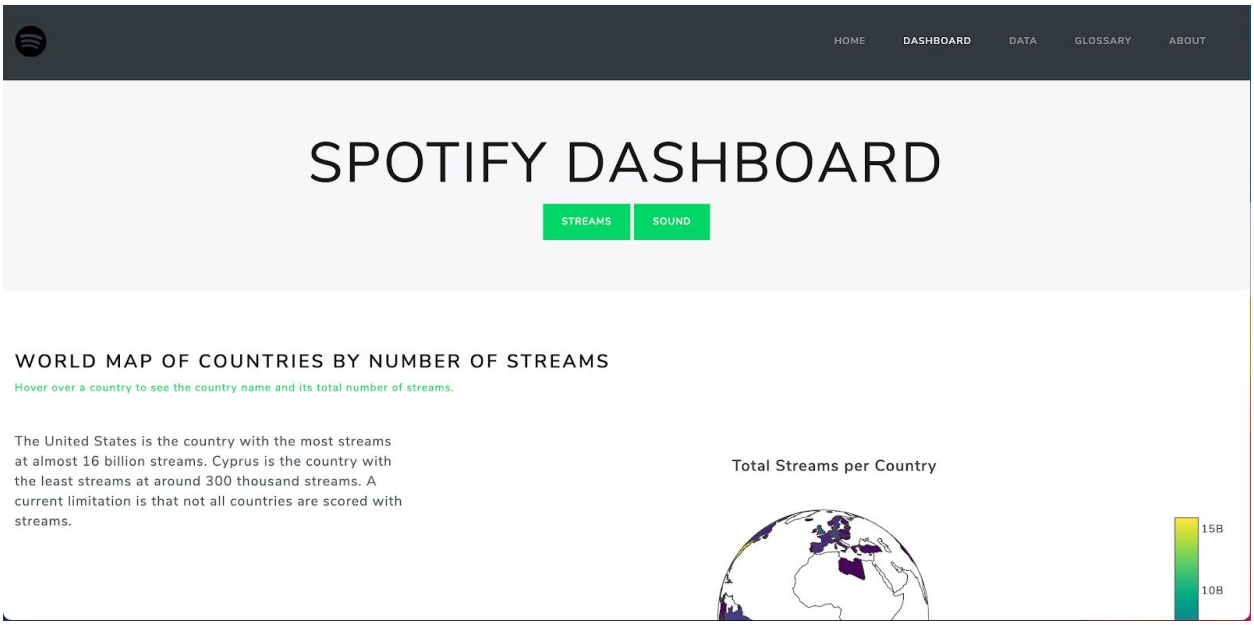
Appendix:

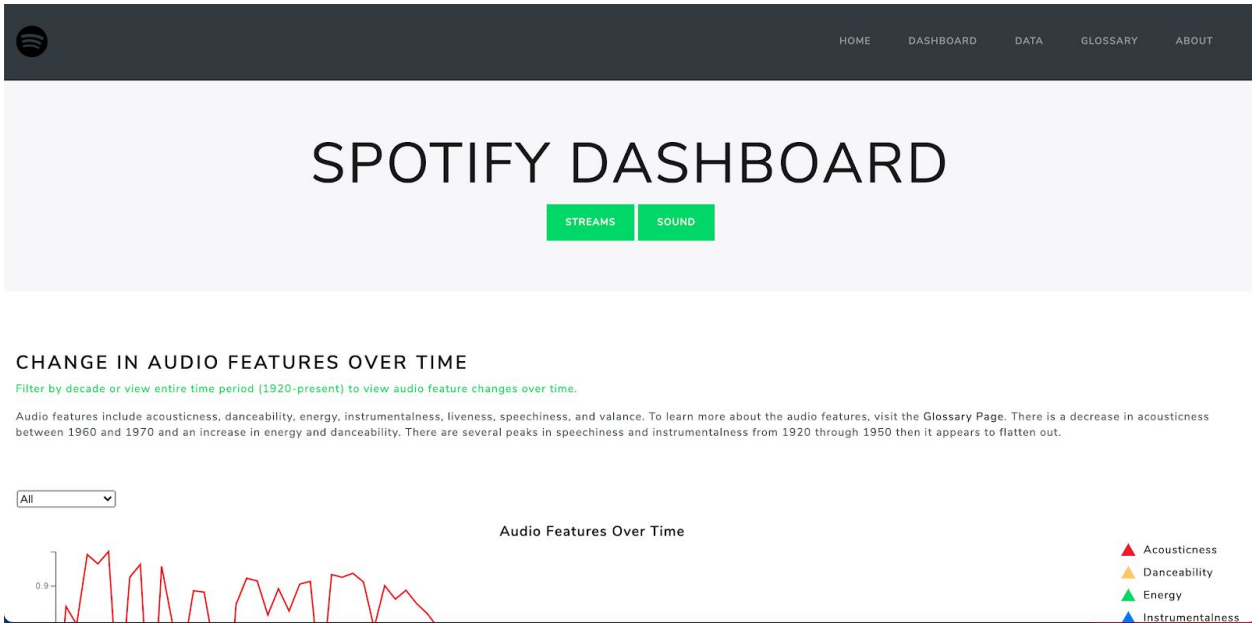
Website:

https://alexarnold630.github.io/SMU_Project2/Spotify/index.html



Dashboards:




HOME
DASHBOARD
DATA
GLOSSARY
ABOUT
SPOTIFY DASHBOARD
STREAMS
SOUND
CHANGE IN AUDIO FEATURES OVER TIME
Filter by decade or view entire time period (1920-present) to view audio feature changes over time.
Audio features include acousticness, danceability, energy, instrumentalness, liveness, speechiness, and valance. To learn more about the audio features, visit the Glossary Page. There is a decrease in acousticness between 1960 and 1970 and an increase in energy and danceability. There are several peaks in speechiness and instrumentalness from 1920 through 1950 then it appears to flatten out.
All
Audio Features Over Time
0.9
Acousticness
Danceability
Energy
Instrumentalness

References

- **Audio Features:** https://developer.spotify.com/documentation/web-api/reference/#endpoint-get-audio-features
- **Data:**
    - **Countries:** https://raw.githubusercontent.com/plotly/datasets/master/2014_world_gdp_with_codes.csv
    - **Kaggle**: https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks/notebooks
    - **Tableau Streams:** https://public.tableau.com/profile/glenn.cable#!/vizhome/Book5_1502/worldmap
- **Inspiration:**
    - https://newsroom.spotify.com/2020-12-01/6-new-features-to-unwrap-in-your-spotify-2020-wrapped/
    - https://public.tableau.com/en-us/gallery/most-successful-songs-2020?tab=viz-of-the-day&type=viz-of-the-day
    - https://graciousquotes.com/music-quotes/
    - https://public.tableau.com/profile/glenn.cable#!/vizhome/Book5_1502/worldmap
- **Spotify logo:** https://www.pngegg.com/en/png-xsvpm
- **Visualizations:**
    - https://plotly.com/javascript/
    - https://d3-legend.susielu.com/